A quantum leap in understanding continuous speech
Webb, Andy

ABSTRACT: E*Trade has purchased software from InterVoice that will enable
its customers to obtain stock quotes by speaking natural language commands.
InterVoice is branching out into banking and brokerage and investment
management markets. In October 1997, Charles Schwab & Co. began using
Nuance Communications' advanced speech recognition technology for its new
mutual funds trading service. A development in the technology that has
enhanced its practicality is the ability to recognize continuous, or
natural, speech. **IBM** 's **ViaVoice** and Dragon Systems' Naturally Speaking
are the first continuous speech recognition products available for the
retail user.   **IBM** 's research facility in Yorktown Heights, New York, is
developing a packaged version of **ViaVoice** for mutual fund stock trading
and online banking with business partners and systems integrators. Accurate
recognition is only likely to be achieved at speeds of between 140 and 160
words per minute with clear enunciation. Although Naturally Speaking is far
ahead of old robot speech systems, it is still slow going in the early
stages.

TEXT: Headnote:

NEW PRODUCTS BY DRAGON SYSTEMS AND **IBM** GEAR VOICE RECOGNITION TECHNOLOGY
UP FOR THE NEXT GENERATION OF USABILITY.

Computer voice recognition used to be something people chuckled over when
watching 1950s science fiction "B" movies. They were prepared to believe
the technology was one day going to happen, but probably not in the format
conjured up by a Brate movie props budget. While the last decade has seen
significant advances in the technology, only recently has it made the leap
into financial markets.

On January 6, InterVoice, a global provider of automated call processing
solutions based in Dallas, announced the receipt of an order from E*Trade,
the Palo Alto-based Internet broker and online investing service, worth an
estimated $3 million, that will enable E*Trade customers to obtain stock
quotes by simply speaking natural language commands. This new order is in
addition to the order E*Trade signed in October to purchase several of
InterVoice's One Voice intelligent software agent platforms, including
software and associated hardware.

"InterVoice's technology is cutting edge," says David R. Ewing, E*Trade's
CIO and senior vice president of technology. Intervoice, which previously
focused on solving repetitive processes like benefits enrollment, job
postings and help desk support, is branching out into banking, and
brokerage and investment management markets -- such as stock quotes/trading
and 401(k) administration. While E*Trade is initially using the technology
to obtain stock quotes, "quickly, easily and with complete
confidentiality," says Ewing, rival Charles Schwab & Co. has already
applied natural language commands to trading.

Last October, Charles Schwab began using Nuance Communications' advanced
speech recognition technology for its new mutual funds trading service. The
system allows customers to simply dial up, state the name of the desired
mutual fund - which automatically triggers a quotation - and then either
make a trade or select a new fund.

Nuance (headquarters in Menlo Park, Calif.) also had previously established
its credentials with Schwab's VoiceBroker, which incorporated Nuance's
Speech Recognition System with its U.S. Stock and Mutual Funds Grammar

System that recognizes over 13,000 stocks, mutual funds and market indexes, and provide options quotes on equities.

Earlier that year, Nuance also announced that U.K. banking group Lloyds TSB would be using the U.K. English version of Nuance6 as part of its telephone banking service that currently handles some 60,000 calls per day.

Not all speech applications involve integrated voice response (IVR), however. Other products, such as Ume Voice, based in Novato, Calif., specialize in making core-recognition technology [middleware] available as class libraries that clients can use to build their own proprietary trading applications. For instance, several brokerage firms use Ume Voice for order entry of mortgage backed securities trades as well as position management in capital markets and information broadcast in currencies.

Because the Wall Street purpose is to achieve productivity gains, this sort of approach allows customers to make existing applications "voice responsive" for any desired commands.

(Illustration Omitted)

Captioned as: **IBM** 's **ViaVoice** uses enrolling speech information to provide regular reminders of how to improve voice recognition.

A development in the technology that has dramatically enhanced its practicality is the ability to recognize continuous, or "natural," speech. Originally, voice recognition systems required the speaker to leave clear gaps between individual words in order to achieve an even remotely respectable recognition success rate. While acceptable for simple menu commands such as "File Open," this "robot speech" requirement made continuous dictation of things such as reports impractical and (after allowing for any necessary corrections) little or no faster overall than having typed them in the first place.

Since 1993, systems integrator Ficomp Systems Inc. of Dayton, N.J., has been perfecting a continuous speech recognition system, that became one of the first voice-activated systems to go live on Wall Street. In June 1996, Bear Stearns implemented Ficomp on its equity trading floor, citing 99.9 percent accuracy and productivity gains. Bear uses The Interpreter 6000 in a Windows-based application it now offers to 450 correspondent traders. In addition, the Chicago Mercantile Exchange developed a speech interface for its price reporting system so that price reporters can speak prices into the system without leaving the trading crowd. At the Bank of Montreal, Interpreter 6000 is taking the place of keyboards in foreign exchange trading.

While the financial applications mentioned so far have been adapted to financial institutions, the two products reviewed here - **IBM** 's **ViaVoice** and Dragon Systems' Naturally Speaking - are the first continuous speech recognition products available for the retail user. (Philips already has such a product, but it runs over a network and is aimed at medium to large corporations). To date, these systems have primarily been used for dictation, for instance, in the medical field. As a result, if you're preparing a report on a merger or acquisition too sensitive for your PA to know about, you no longer have to struggle with your rusty typing skills or imitate speech patterns of the creature from the planet Zog!

There are indications, however, that these dictation systems will be applied to financial services in the future. For instance, Ume Voice has a copy of **ViaVoice** . **IBM** 's research facility in Yorktown Heights, N.Y., is developing a packaged version of **ViaVoice** for mutual fund stock trading and online banking with business partners and systems integrators, that it expects to release later this year, according to an **IBM** spokeswoman. She adds that **ViaVoice** is a technology that is available for large vocabulary applications in the business market.
However, while "continuous" speech recognition does mean that you can speak without pausing between words it does not mean (at least to start with) that you can speak in your normal everyday manner. Accurate recognition is only likely to be achieved at speeds of between 140 and 160 words per

minute with clear enunciation. While the gaps may no longer be required, some delineation between words is still necessary. The "runtogether" way in which most people speak is still something of a Holy Grail for speech recognition software in general.

In order to improve recognition, both products use **artificial intelligence** techniques to analyze and learn from a user's individual speech patterns and voice. As a result, neither comes cheap in terms of system requirements.

Dragon recommends a minimum machine spec for NaturallySpeaking of 32MB of RAM for Windows 95 operation (48MB under Windows NT) and a P133 processor. **IBM** recommends an identical minimum memory spec for **ViaVoice** but with a P166 processor.

The reviews were conducted on a PC running Windows NT with a Pentium Pro 200 processor and 128MB RAM - even with that horsepower, recognition wasn't instantaneous with either product.
(Photograph Omitted)

Captioned as: While the uses and demands for usable voice recognition technology are growing, mastering the "runtogether" way in which people speak remains the industry's Holy Grail.

COMPANY NAMES:
E-Trade Group Inc
InterVoice Inc (DUNS:11-743-9273)
Charles Schwab & Co Inc (DUNS:06-302-0804 TICKER:SCH)
Nuance Communications
Dragon Systems Inc
GEOGRAPHIC NAMES: US

DESCRIPTORS: Voice recognition; Interactive computer systems; Technological
    change; Manycompanies; Manyproducts; Financial services
CLASSIFICATION CODES: 9190 (CN=United States); 5240 (CN=Software & systems)
    ; 8130 (CN=Investment services)
?

- Re. ViaVoice: It's clear to me that this is an AI application. In addition to the "learning" mentioned in the patent, see attached Dialog article "A Quantum Leap in Understanding Continuous Speech", Spring 1998 which mentions that it includes AI. I found others in Dialog that made further comments about ViaVoice and AI with respect to semantics/grammar, not just speech training.

-

- Just focusing on claim 20, I think Pickering does have an interactive terminal that elicits verbal instructions from the customer and semantically processes the verbal instructions via the AI routines to parse the instruction. However, although Pickering does say that there are additional error processing steps including transferring the call to a human operator e.g. if the audio quality is bad if no text is produced, this is not necessarily nor inherently the result of a problem in the semantic processing, nor is this something that is necessarily handled by the AI routines per se. The additional error processing steps are not directly tied to failure of semantic processing. Even though semantic processing would certainly fail when audio quality is bad, it is not inherent that the error processing steps taught by Pickering are incorporated as part of the semantic processing. The error steps of Pickering could equally be done in other steps outside the AI routines to merely detect connection quality, detect noise levels, etc. This is not to say that the claim is patentable, just that Pickering would not be enough for a 102.

9/28/04